Near Infrared Face Based Biometric Key Binding

Meng Ao and Stan Z. Li*

Center for Biometrics and Security Research,
Institute of Automation, Chinese Academy of Sciences,
95 Zhongguancun Donglu, Beijing 100190, China
{mao,szli}@cbsr.ia.ac.cn

Abstract. Biometric encryption is the basis for biometric template protection and information security. While existing methods are based on iris or fingerprint modality, face has so far been considered not reliable enough to meet the requirement for error correcting ability. In this paper, we present a novel biometric key binding method based on near infrared (NIR) face biometric. An enhanced BioHash algorithm is developed by imposing an NXOR mask onto the input to the subsequent error correcting code (ECC). This way, when combined with ECC and NIR face features, it enables reliable binding of face biometric features and the biometric key. Its ability for template protection and information cryptography is guarantied by the theory of encryption. The security level of NIR face recognition system is thereby improved. Experimental results show that the security benefit is gained with a sacrifice of 1-2% drop in the recognition performance.

Keywords: BioHash, ECC, Key Binding, NIR Face, Biometric Encryption.

1 Introduction

Cryptology of security keys plays an important role in information security [14]. When the key and its owner are separated, the computer know whether the key is correct but not the user. If an attacker gets the key, the information will be decrypted easily. This presents a problem in security.

A biometric based system gets personal identity, but relies on the operating system or access key to control access to the information stored on the system. So, biometrics alone can solve the security problem only partially.

In addition, biometric has its inherent shortcomings. Biometric characteristics (especially physiological biometrics) are largely immutable, resulting in permanent biometric compromise when stolen or leaked. Therefore, how to protect biometric templates becomes an important issue. In traditional systems using passwords, Message-Digest Algorithm 5 (MD5) is widely used to store passwords

* Corresponding author.

M. Tistarelli and M.S. Nixon (Eds.): ICB 2009, LNCS 5558, pp. 376–385, 2009.

[13]. The system only stores the hashed password, which ensures the system will not leak the password.

A good solution to these problems is to bind the access key with its owner's identity for each piece of restricted information. There are two approaches to biometric based information security. One is to convert a biometric template into a stable vector and use MD5 or other one-way algorithm to protect the biometric template. A problem with this approach is that the biometric template is still unchangeable. The other is to bind a biometric feature to a key. If the system is cracked by brutal force attack, the system administrator just needs to change a binding key.

However, there is a problem in applying biometrics to security key generation: Biometric features are subject to certain degree of variation and so may not be 100% the same for all feature extraction session. To tackle such inherent variability of biometrics, error correcting code (ECC) is used to correct errors or noise in biometric features, to finally enhance the reliability of biometric key binding [12,29,28,27]. Since the ECC algorithms only work in binary system with hamming distance. In order to use ECC algorithms, BioHash[25,23,26,32,31] method can be introduced to convert the biometric feature vectors into binary strings. Therefore, BioHash and ECC theory can be combined to enhance key binding to a biometric system.

Most of existing works in this direction are mainly based on fingerprint and iris modalities [5,10,11,18,9,20,4,16,3,2,19,7]. This is because fingerprint and iris features are more stable and can be easily incorporated with ECC [8,24,30,1].

There has been some literature reporting key binding with face biometric features[17,15]. These works are based on a visible light face recognition. The experiments in these papers are close-set face recognition. That means these works can not be used in real application. The main reason is that face has thus far been considered not stable enough to be error-corrected by ECC. Face biometric is affected by lighting condition, expression, and other factors [34]. The variability in face features often exceeds the correcting ability of ECC.

Near infrared (NIR) face recognition introduced recently has achieved great success. Not only it overcomes the illumination problem but also achieved significant higher accuracy performance than conventional visual face biometrics [21]. It can be used reliably for 1-to-many face identification. This new technology makes face biometric key binding possible.

In this paper, we present a novel biometric key binding method based on near infrared (NIR) face biometric. The contributions are the following: First, we develop an enhanced BioHash algorithm by imposing an NXOR mask onto the input to the subsequent error correcting code (ECC). Second, we present and NIR face based key binding for improving the security level of NIR face recognition system. the BioHash is combined with ECC and NIR face features to enables reliable binding of face biometric features and the biometric key.

While the ability of the BioHash algorithm for template protection and information cryptography is guarantied by the theory of encryption, we present

experimental results show that the security benefit is gained with a sacrifice of 2% drop in the recognition performance.

The rest of the paper is organized as follows. Section 2 introduces BioHash method and the enhanced BioHash. Section 3 describes the NIR face based encryption algorithm. Section 4 shows the experiment result.

2 Enhanced BioHash

BioHash represents a series of operations which combine a high dimensional feature vector with a user-specific tokenised random numbers (TRN) to produce a binary bit string, so that the similarity score between two feature vectors can be measured by the hamming distance [26,32,6]. The basic idea of BioHash is to use threshold calculation to obtain a bit 0 or 1. The following describe three BioHash procedures for converting a feature vector to a binary string.

BioHash 1. By repeatedly converting a feature vector to a binary string by comparing it with a random vector with same length.

1. Use a feature extraction technique to extract the biometric feature. The biometric feature is represented in a vector form, $x \in \mathscr{R}^n$, with n denoting the feature length of x. Repeating x m times to obtain a new feature vector $X = \{x, ..., x\}$.
2. Use token to generate a pseudo-random vectors $r \in \mathscr{R}^{n \times m}$ based on a seed. m is a integer. The entire distribution of r is same as x.
3. Use the following threshold calculation to obtain a binary string $b_1 b_2 ... b_{n \times m}$ with length $n \times m$

$$b_i = \begin{cases} 1 \ if \ X_i > r_i \\ 0 \ if \ X_i \leqslant r_i \end{cases} i = \{1, ..., n \times m\} \tag{1}$$

The n-dimensional biometric feature x is thus converted to a binary string b of length $n \times \ m$.

BioHash 2. By calculating the dot product of the feature vector and several random vectors, and then comparing the dot product results with a threshold.

1. Employ the input token to generate a set of pseudo-random vectors, $r_i \in \mathscr{R}^M$ for $i = 1, \ldots, m$ based on a seed.
2. Apply the Gram-Schmidt process to $r_i \in \mathscr{R}^M$ for $i = 1, \ldots, m$ to obtain a set of orthonormal vectors $p_i \in \mathscr{R}^M | i = 1, ..., m$, also called Tokenised Random Number, TRN.
3. Calculate the dot product of v, the feature vector obtained from first step and each orthnonormal vector in TRN, p_i, such that $\langle v, p \rangle$.
4. Use a threshold τ to obtain a binary string, b whose elements are defined as

$$b_i = \begin{cases} 1 \ if \ \langle v, p \rangle > \tau \\ 0 \ if \ \langle v, p \rangle \leqslant \tau \end{cases} i = \{1, ..., m\} \tag{2}$$

BioHash 3. Converting a feature vector to a binary string by calculating the range of arguments of the complex, which is generated by adding a random imaginary to the feature vector.

1. Use token to generate a set of pseudo-random vectors, $r_i \in \mathscr{R}^M | i = 1, ..., m$ with the entires distributed according to $N(0, 1)$ and apply the Gram-Schmidt process to transform the basis $r_i \in \mathscr{R}^n | i = 1, ..., m$ into an orthonormal set of $r, r_{\perp i} \in \mathscr{R}^n | i = 1, ..., m$.
2. Mix the x with $r_{\perp i}$ iteratively to form the complex number, $\{z_i = x_i + r_{\perp i} j | i = 2, ..., n\}$, where $j = \sqrt{-1}$ and calculate their complex arguments, $\{arg(z_i) \in \mathscr{R}^n | i = 1, ..., n\}$.
3. Average the complex arguments, $\left\{\alpha_i = \frac{1}{n} \sum_{j=1}^{n} arg(z_i) \in \mathscr{R}^m | i = 1, ..., m\right\}$ where $-\pi \leqslant \alpha_i < \pi$ and $m < n$. Then we get the binary string b

$$b_i = \begin{cases} 1 \; if \; -\pi < \alpha < 0 \\ 0 \; if \; 0 \leqslant \alpha < \pi \end{cases} i = \{1, ..., m\} \tag{3}$$

Enhanced BioHash

Among the three BioHash algorithms, BioHash 1 is advantageous in that binary strings it generates is much longer than the others and easy to calculate. With a longer binary string, we can bind into longer key, which increases the security of system.

However, this incurs a problem that we cannot using an ECC algorithm to correct the binary string: The binary string obtained can be any one in the space $\{0, 1\}^n$; however, ECC cannot decode some of these. ECC coding and decoding algorithms are generally used in pair. There are some strings in the space $\{0, 1\}^n$ which are out of ECC decoding space.

Here we present an enhanced BioHash method to solve this problem. Suppose that b is a binary string with length n which is converted from a biometric feature vector. For any binary string b' of length n, we generate a mask string $M = b \; NXOR \; b')$ by

$$M_i = b_i \; NXOR \; b'_i, i = 1, ..., n \tag{4}$$

Here NXOR (Not eXclusive OR) is a logical operation. The NXOR is defined as follows

$$1 \; NXOR \; 1 = 1, 0 \; NXOR \; 0 = 1, 1 \; NXOR \; 0 = 0, 0 \; NXOR \; 1 = 0. \tag{5}$$

An NXOR between two strings is defined as NXOR on each bit respectively. The binary string b is then converted to b' using the operation $b' = b \; NXOR \; M$ by

$$b_i \; NXOR \; M_i = b'_i, i = 1, ..., n \tag{6}$$

With the same mask M, the hamming distance between b_1 and b_2 does not change because the property of NXOR operator, that is,

$$D_{Hamming}(b_1, b_2) = D_{Hamming}((b_1 \; NXOR \; M), (b_2 \; NXOR \; M)) \tag{7}$$

3 NIR Face Based Key Binding

The motivation of NIR face based key binding is the following: We need a face encryption system. Error correcting code (ECC) is usually needed to solve the problem incurred by variation in biometric features. The ECC based method requires that the biometric module be able to achieve a minimum performance that is associated with the ability ranger of ECC. The stability of NIR face features can satisfy this requirement of ECC, as will be explained.

When the key binding method is introduced into a face recognition system, the enrollment and recognition processes are both changed. So, a description of face enrollment and recognition for biometric key binding will be provided. How to use multi-image for enrollment to improve the system performance will be described. The security level of the key binding NIR face recognition system and the external function of the system will be discussed, too.

3.1 Error Correcting Code

Error Correcting Code is a code in which each data signal conforms to specific rules of construction so that departures from this construction in the received signal can generally be automatically detected and corrected. The basic idea is for the transmitter to apply one or more of the above error detecting codes; then the receiver uses those codes to narrow down exactly where in the message the error (if any) is. Shannon's theorem points out that the maximum attainable efficiency of an error-correcting scheme versus the levels of noise interference expected.

Two main categories are convolutional codes and block codes. Examples of the latter are Hamming code, BCH (Bose-Chaudhuri-Hocquenghem) code, Reed-Solomon code, Reed-Muller code, Binary Golay code, and low-density parity-check codes [33]. Here we use BCH code to convert the unstable BioHash binary string to a stable binary string.

In the error correcting theory the ratio of error bits to total bits has an upper limit. Suppose b_1 and b_2 are the two BioHash strings from the same person. If we use the error correcting code to transform the BioHash binary string to a stable binary string, the difference of two binary strings from the same person should be less than the upper limit. As we know, face images usually vary large due to various lighting changes which may make the difference of features from the same person larger than the upper limit. To deal with this problem, here we choose NIR face recognition as the basic system which has an advantage on face encryption. See Fig.1.

The use of NIR imaging brings a new dimension for applications of invisible lights for face detection and recognition [21]. It not only provides appropriate active frontal lighting but also minimizes lightings from other sources. The following figure compares NIR face images with visible light (VL) face images. The VL images contain large performance-deteriorating lighting changes whereas the NIR are good for face recognition. The fixed lighting direction much simplifies the problem of face recognition. Face recognition algorithms are more stable in under controlled lighting. That meets the requirement of error correcting code theory.

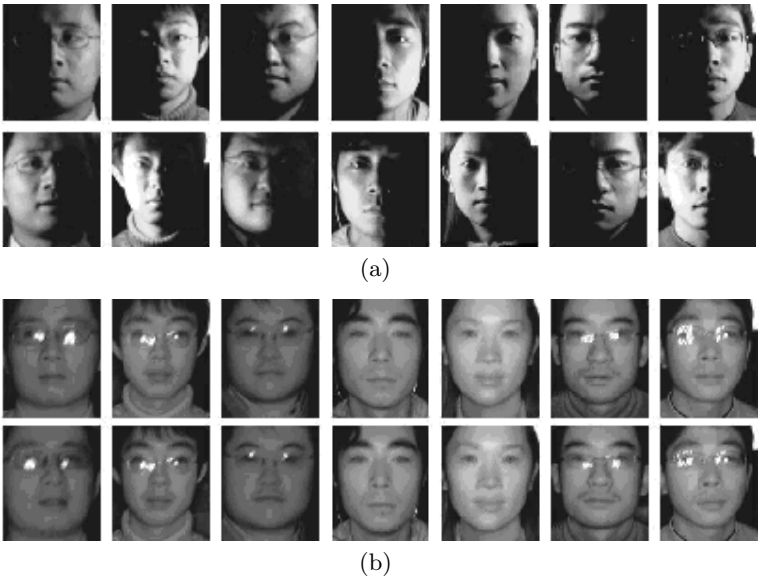


Fig. 1. Comparison of VL images (a) vs. NIR images (b) taken in the same environmental lighting conditions

3.2 Face Enrollment

The enrollment process is showed in Fig.2(a). Suppose the length of face feature is n. The length of BioHash binary string is $n \times m$. The threshold of the hamming distance is τ. The the enroll process is:

1. Face feature extraction. Extract a feature vector x from a NIR face image.
2. Use the basic BioHash method to convert the face feature vector x to a binary string b. The seed of the pseudo-random vector in BioHash is s.
3. Generate a random integer vector. Convert the random vector to binary system. Here we get a random binary string k with the length of t. This is the binded key.
4. Use BCH coding algorithm to transform k to k_{BCH}. The length l of k_{BCH} is a little smaller than $n \times m$ or equal. Use the first l bits of b to get the mask $M = k_{BCH}\ NXOR\ b$.
5. Record the M , s and k values in the database. k should be protected by MD5 or other one-way algorithm.

Item 5 refers the storage in the database. It is obvious that we can no nothing about the face feature just from M and s. However, if k can be obtained by the attacker, the attacker may concoct a new feature vector which can pass the system. So the binded key k should be protected by MD5.

Note that in item 1 in BioHash 1, the face feature vector x is repeated m times. This brings multiimages for the enrolled person to improve the performance of

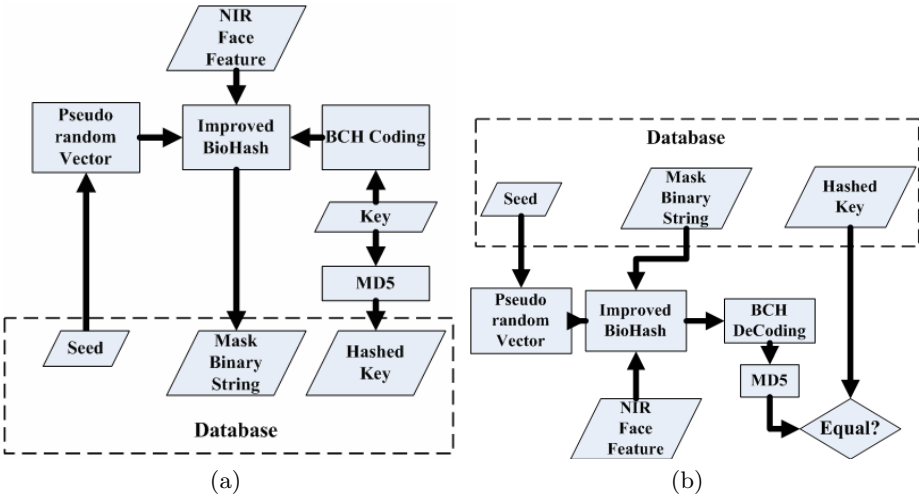


Fig. 2. (a) The enrollment process. (b) The recognition process.

the system. Suppose there are several images from one person with corresponding feature vectors $x_1, x_2, ..., x_s$. We use $\{x_1, x_2, ..., x_s, ..., x_1, x_2, ..., x_s\}$ instead of $\{x, x, ..., x\}$ in item 1.

3.3 Face Recognition

The recognition, shown in Fig.2(b), is described as follows:

1. Face feature extraction. Extract a feature vector x' from a NIR face image.
2. Get s from the database. Use s to generate a seudo-random vector for BioHash. Convert x to the binary string b'.
3. Get M from the database. Calculate the masked string $k'_{BCH} = b' \, NXOR \, M$.
4. Use BCH decoding algorithm to transform k'_{BCH} to k'.
5. Check if k' is k by MD5. if $k' = k$, then x' and x are from the same person.

Because the probability of collision of MD5 is very small, the performance of the recognition mainly depends on the result of BioHash.

We can see in enrollment and recognition processes that the face template stored in the system are protected by MD5 or other one-way algorithm. Here the template protection depends on the security of encryption algorithms. The security of the face recognition system is guaranteed by the encryption theory [13].

The NIR face base key binding can be used in information cryptography. The binding key can be the same key for encryption and decryption of a message. This application has a wide range of prospects.

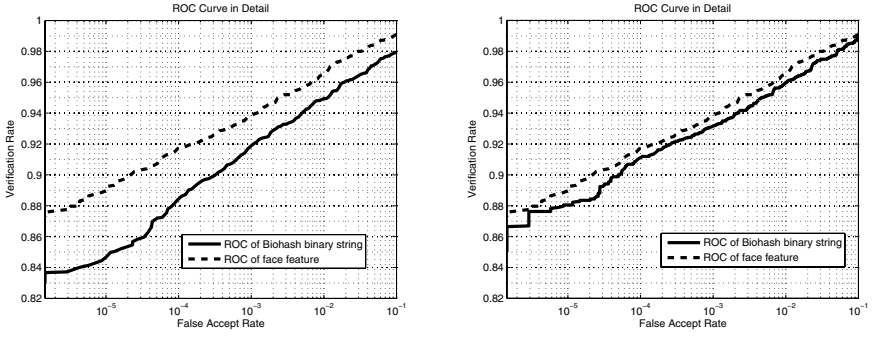
4 Experiments

Experiments are done to compare the performance of the original NIR face features algorithm and the enhanced BioHash binary strings with face key binding

incorporated. The purpose is to evaluate how much sacrifice in recognition accuracy the system has to pay for the gained security.

The original algorithm converts a 142×120 face image to a 256-dimensional feature vector by the algorithm in paper [22]. There are no intersection between the training database and testing database. The testing database contains 1176 images, which are from 294 people with 4 images per person. Every image is compared with other images. So the number of similarity scores is $C_{1176}^2 = 690900$.

The original system uses $L2$ distance as the similarity score. The key banding system checks whether the result of MD5 is equal. BCH decode algorithm ensures the two binary strings getting different codes when the hamming distance is bigger than BCH threshold. So we only need to calculate ROC of the BioHash binary string with hamming distance. The result is shown in Fig.3(a).



(a) (b)

Fig. 3. The comparation of ROC curve between BioHash binary string and raw face feature: (a) When Biohash binary string length is 4095, and (b) When Biohash binary string length is 66175

We set the length of BioHash binary string to be 4095, threshold 285, and corresponding key length as 707. The enhanced BioHash binary string achieves verification rate (VR) of 92.01% at FAR = 0.11%. This decreased VR by 2% from the original system. However, this method offers security to the face recognition system, and information protection ability for the face recognition system.

In BCH algorithm, when the total number is 4095, the corresponding error bits can be 1 to 343. That means in our method, the range of FAR is from 0 to 8.32%, and the range of VR is from 0 to 97.79%.

The BCH code length can be longer in order to make a long key. When the code length increases, BCH algorithm takes more time to code and decode. That also makes the brute force search very difficult. See Fig.3(b). When we set the length of BioHash binary string to be 66175, threshold 3407, and corresponding key length as 11131. And with the long biohash binary string, the VR also increases to 93.15%.

5 Conclusion

This paper has presented a novel biometric key binding method, enhanced Bio-Hash. The method, when incoporated with NIR face biometric, enables face biometric based template protection, file encrypting and many other applications. The gain in security is proved by existing theory, with a little drop in recognition accuracy.

Acknowledgements. This work was supported by the following funding resources: National Natural Science Foundation Project #60518002, National Science and Technology Support Program Project #2006BAK08B06, National Hi-Tech (863) Program Projects #2006AA01Z192, #2006AA01Z193, and #2008AA01Z124, Chinese Academy of Sciences 100 people project, and AuthenMetric R&D Funds.

References

1. Alabbadi, M., Wicker, S.B.: A digital signature scheme based on linear error-correcting block codes (1995)
2. Bakhtiari, S., Pieprzyk, J.: On the weakness of gongs collisionful hash function. Journal of Universal Computer Science 3, 185–196 (1997)
3. Bakhtiari, S., Safavi-naini, R., Pieprzyk, J.: On password-based authenticated key exchange using collisionful hash functions (1996)
4. Berson, T., Gong, L., Lomas, T.: Secure, keyed, and collisionful hash functions. December 1993. Included in Technical Report SRI-CSL-94-08, Computer Science Laboratory, SRI International, Menlo Park, California (May 1994)
5. Bodo, A.: Method for producing a digital signature with aid of a biometric feature, German Patent, DE 4243908A1 (1994)
6. Cheung, K.H., Kong, A.W.-K., Zhang, D., Kamel, M., You, J.: Revealing the secret of facehashing. In: Zhang, D., Jain, A.K. (eds.) ICB 2006. LNCS, vol. 3832, pp. 106–112. Springer, Heidelberg (2006)
7. Clancy, T.C., Kiyavash, N., Lin, D.J.: Secure smartcardbased fingerprint authentication (2003)
8. Crepeau, C., Kilian, J.: Achieving oblivious transfer using weakened security assumptions (1988) (extended abstract)
9. Daugman, J.: How iris recognition works. IEEE Transactions on Circuits and Systems for Video Technology 14, 21–30 (2004)
10. Davida, G., Frankel, Y., Matt, B.: On enabling secure applications through off-line biometric identification. In: 1998 IEEE Symposium on Security and Privacy, Proceedings, May 1998, pp. 148–157 (1998)
11. Davida, G.I., Frankel, Y., Matt, B.J., Peralta, R.: On the relation of error correction and cryptography to an offline biometric based identification scheme (1999)
12. Dodis, Y., Ostrovsky, R., Reyzin, L., Smith, A.: Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. SIAM J. Comput. 38(1), 97–139 (2008)
13. Feldmeier, D.C., Karn, P.R.: Unix password security - ten years later (1990)
14. Garrett, P.B.: Making, Breaking Codes: Introduction to Cryptology. Prentice Hall PTR, Upper Saddle River (2000)

15. Goh, A., Ling, D.N.C., Goh, A.: Computation of cryptographic keys from face biometrics. In: Lioy, A., Mazzocchi, D. (eds.) CMS 2003. LNCS, vol. 2828, pp. 1–13. Springer, Heidelberg (2003)
16. Gong, L.: Collisionful keyed hash functions with selectable collisions. Inf. Process. Lett. 55(3), 167–170 (1995)
17. Jin, A.T.B., Ling, D.N.C., Goh, A.: Personalised cryptographic key generation based on facehashing. Computers & Security 23(7), 606–614 (2004)
18. Juels, A.: A fuzzy vault scheme (2002)
19. Juels, A.: Fuzzy vaults: Toward secure client-side matching (2002)
20. Juels, A., Wattenberg, M.: A fuzzy commitment scheme (1999)
21. Li, S., Chu, R., Liao, S., Zhang, L.: Illumination invariant face recognition using near-infrared images. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(4), 627–639 (2007)
22. Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S.Z.: Learning multi-scale block local binary patterns for face recognition. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 828–837. Springer, Heidelberg (2007)
23. Lumini, A., Nanni, L.: An improved biohashing for human authentication. Pattern Recogn. 40(3), 1057–1065 (2007)
24. McEliece, R.J.: A public-key cryptosystem based on algebraic coding theory. Deep Space Network Progress Report 44, 114–116 (1978)
25. Nanni, L., Lumini, A.: Empirical tests on biohashing. Neurocomputing 69(16-18), 2390–2395 (2006)
26. Nanni, L., Lumini, A.: Random subspace for an improved biohashing for face authentication. Pattern Recogn. Lett. 29(3), 295–300 (2008)
27. Nichols, R.K.: ICSA Guide to Cryptography. McGraw-Hill Professional, New York (1998)
28. Soutar, C., Roberge, D., Stoianov, A., Gilroy, R., Kumar, B.V.: Biometric encryption: enrollment and verification procedures, vol. 3386, pp. 24–35. SPIE (1998)
29. Soutar, C., Roberge, D., Stoianov, A., Gilroy, R., Vijaya Kumar, B.V.K.: Biometric Encryption using image processing (April 1998)
30. Stern, J.: A new identification scheme based on syndrome decoding (1994)
31. Teoh, A., Jin, B., Connie, T., Ngo, D., Ling, C.: Remarks on biohash and its mathematical foundation. Inf. Process. Lett. 100(4), 145–150 (2006)
32. Teoh, A.B.J., Kuan, Y.W., Lee, S.: Cancellable biometrics and annotations on biohash. Pattern Recogn. 41(6), 2034–2044 (2008)
33. Wikipedia. Error detection and correction — Wikipedia, the free encyclopedia (2008) (Online; accessed November 6, 2008)
34. Zhao, W., Chellappa, R., Phillips, P., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Surveys, 399–458 (2003)